## Supplemental Materials for

# Political Plausible Deniability: Political Difference can Divert Attributions of Socially Unacceptable Bias

**This PDF file includes:**

**Supplemental Material A: Studies 3a-c Minor Deviation from Preregistration**

In our preregistration, we planned to use a scenario for Studies 3a-c that described the actor as a coworker (rather than a manager) declining to help the target (another coworker). However, we later became concerned that this design would conflate two different dimensions across studies: the stakes/conflict type involved and the status of the actor versus the target. Our intention in designing Studies 2a-c, 3a-c, and 4a-c was to juxtapose different stakes involved in a conflict: low stakes (declining to help), medium stakes (expressing anger), and high stakes (declining a promotion). We did not intend to conflate that dimension with the status of the actor versus the target. Therefore, we revised the scenarios in Studies 3a-c to describe a manager (rather than a coworker) declining to help an employee.

**Supplemental Material B: Pretest 1a Coding Criteria for Socially Acceptable Bias**

1) **Consensus about acceptability**
   - Majority, societal norms, cultural practices
   - References to "most people," "society," "population"
   - Mentions of norms, traditions, or ingrained practices
   - Descriptions of behaviors being "common," "widely accepted"
   - **Examples**:
     - *Socially acceptable means the behavior is accepted by the general population*
     - *Certain forms of bias are perpetuated by cultural norms and stereotypes deeply ingrained in society*

2) **It's okay to express it publicly**
   - The person expressing the bias feels comfortable doing so
   - Idea that a bias is acceptable if it can be openly expressed without discomfort
   - References to public expression or appropriateness in social settings
   - Mentions of comfort or confidence in expressing behaviors or biases
   - **Examples**:
     - *Socially acceptable refers to actions that are accepted and viewed as appropriate to engage in with other people*
     - *I think of socially acceptable as something people are just used to hearing*

3) **No repercussions for expressing it**
   - Mentions that social acceptability is tied to the absence of penalties, backlash, or negative consequences
   - Explicit mention of consequences (or their absence)
   - References to punishment, ostracism, or social judgment
   - **Examples**:
     - *Socially acceptable could mean that society at large will not punish (via law, social ostracizing, etc.) said behavior*
     - *You will not suffer any social consequences for the behavior, like losing friends or putting your job at risk*

**Distinguishing between the three codes:**

- 'Consensus' reflects **collective agreement** about norms and standards
  - Focuses on what **society** collectively deems acceptable
  - If a response mentions societal norms or collective agreement, it likely reflects the 'consensus' code
- 'Okay to express in public' focuses on **comfort** with expressing a particular bias
  - Focuses on **individual-level** expression in public settings
  - If the response mentions comfort or appropriateness in public settings, it likely reflects the 'okay to express in public' code
- 'No repercussions' focuses on **consequences and penalties**
  - If the response mentions freedom from punishment or backlash, it likely reflects the 'no repercussions' code

**Supplemental Material C: Pretest 2 Results for All Demographic Characteristics**

To what extent do you think other people would say that the following forms of bias/discrimination are socially acceptable versus socially unacceptable? (1 = *Very Socially Unacceptable*, 7 = *Very Socially Acceptable*)

Bias/discrimination based on...

| | |
|---|---|
| Race: | $M = 2.62$ ($t(199) = 10.63$, $p < .001$) |
| Disability: | $M = 2.63$ ($t(199) = 10.51$, $p < .001$) |
| Sexual orientation: | $M = 2.97$ ($t(199) = 7.98$, $p < .001$) |
| Gender: | $M = 3.17$ ($t(199) = 6.00$, $p < .001$) |
| Religion: | $M = 3.51$ ($t(200) = 3.49$, $p < .001$) |
| Nationality: | $M = 3.52$ ($t(199) = 3.58$, $p < .001$) |
| Body weight: | $M = 3.73$ ($t(200) = 1.93$, $p = .055$) |
| Social class: | $M = 3.89$ ($t(200) = 0.87$, *ns*) |
| Age: | $M = 4.18$ ($t(200) = 1.40$, *ns*) |
| Community type (rural vs. urban): | $M = 4.43$ ($t(200) = 3.47$, $p < .001$) |
| Job type (blue vs. white collar): | $M = 4.43$ ($t(200) = 3.48$, $p < .001$) |
| Education level: | $M = 4.51$ ($t(199) = 4.16$, $p < .001$) |
| Physical attractiveness: | $M = 4.54$ ($t(200) = 3.90$, $p < .001$) |
| Attractiveness: | $M = 4.56$ ($t(200) = 4.03$, $p < .001$) |
| Political party: | $M = 4.79$ ($t(200) = 6.53$, $p < .001$) |

**Supplemental Material D: Survey Items (Studies 1a-c, 2a-c, 3a-c, 4a-c, and 5)**

**Demographics (all studies)**

[Gender] How do you describe your gender?

- Male
- Female
- Non-binary / third gender
- Prefer to self-describe [text box]
- Prefer not to say

[Race] Please select which race / ethnicity you identify as. (Please select all that apply.)

- White / Caucasian
- Black / African American
- Hispanic / Latino
- Asian / Asian American
- Other [text box]

[Attention 1] To help us keep track of who is paying attention, please select any number other than 4.

- 1
- 2
- 3
- 4

[Sexual orientation] How do you describe your sexual orientation?

- Straight
- Gay
- Bisexual
- Other [text box]
- Prefer not to say

[Partisanship] Generally speaking, how do you think about your partisan affiliation? (If you identify with a third party, please indicate whether you tend to lean toward one of the major parties or consider yourself truly independent.)

- I am a strong Republican
- I am a Republican
- I tend to lean towards the Republican Party
- I am a true independent
- I tend to lean towards the Democratic Party
- I am a Democrat
- I am a strong Democrat

[Attention 2] Please write “twenty-five” using numbers. [text box]

**Studies 1a-c** (differences between a/b/c in bold)

[Control condition]
Now read the following scenario carefully: Imagine that one day at work, you witnessed an intense conversation between **[Sam/John/Will]**, a **[White/male/straight]** manager, and **[Alex/Sarah/Mike]**, a **[Black/female/openly gay]** sales associate. Angered by **[Alex/Sarah/Mike]**'s decision to refund money back to a customer, **[Sam/John/Will]** told **[Alex/Sarah/Mike]** that he did not understand "why the firm keeps hiring **[people/women/people]** like you."

[Bleeding heart liberal treatment condition]
Now read the following scenario carefully: Imagine that one day at work, you witnessed an intense conversation between **[Sam/John/Will]**, a **[White/male/straight]** manager, and **[Alex/Sarah/Mike]**, a **[Black/female/openly gay]** sales associate. Angered by **[Alex/Sarah/Mike]**'s decision to refund money back to a customer, **[Sam/John/Will]** told **[Alex/Sarah/Mike]** that he did not understand "why the firm keeps hiring bleeding heart liberal **[people/women/people]** like you."

[Snowflake liberal treatment condition]
Now read the following scenario carefully: Imagine that one day at work, you witnessed an intense conversation between **[Sam/John/Will]**, a **[White/male/straight]** manager, and **[Alex/Sarah/Mike]**, a **[Black/female/openly gay]** sales associate. Angered by **[Alex/Sarah/Mike]**'s decision to refund money back to a customer, **[Sam/John/Will]** told **[Alex/Sarah/Mike]** that he did not understand "why the firm keeps hiring snowflake liberal **[people/women/people]** like you."

[Woke liberal treatment condition]
Now read the following scenario carefully: Imagine that one day at work, you witnessed an intense conversation between **[Sam/John/Will]**, a **[White/male/straight]** manager, and **[Alex/Sarah/Mike]**, a **[Black/female/openly gay]** sales associate. Angered by **[Alex/Sarah/Mike]**'s decision to refund money back to a customer, **[Sam/John/Will]** told **[Alex/Sarah/Mike]** that he did not understand "why the firm keeps hiring woke liberal **[people/women/people]** like you."

[Racism/Sexism/Heterosexism attributions] Why do you think **[Sam/John/Will]** (the manager who expressed anger) said that to **[Alex/Sarah/Mike]**? [Open-ended textbox]

[Manipulation check] To what extent do you think **[Sam/John/Will]** (the manager who expressed anger) was talking about politics when he spoke to **[Alex/Sarah/Mike]**?

- He was definitely talking about politics
- He was probably talking about politics
- I don't know/Not sure
- He was probably not talking about politics
- He was definitely not talking about politics

**Studies 2a-c** (differences between a/b/c in bold)

[Control condition] Same as Studies 1a-c

[Bleeding heart liberal treatment condition] Same as Studies 1a-c

[Snowflake liberal treatment condition] Same as Studies 1a-c

[Woke liberal treatment condition] Same as Studies 1a-c

[Manipulation check] Same as Studies 1a-c

[Political bias] Now, please indicate the extent to which you agree or disagree with each statement about **[Sam/John/Will]** from the scenario you just read. As a reminder, **[Sam/John/Will]** was the manager who expressed anger. [Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*]

- **[Sam/John/Will]** is prejudiced against Democrats.
- **[Sam/John/Will]** is politically biased.
- **[Sam/John/Will]** holds negative stereotypes about Democrats.

[Racism/Sexism/Heterosexism perceptions] Now, please indicate the extent to which you agree or disagree with each statement about **[Sam/John/Will]** from the scenario you just read. As a reminder, **[Sam/John/Will]** was the manager who expressed anger. [Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*]

- **[Sam/John/Will]** is prejudiced against **[Black people/women/gay people]**.
- **[Sam/John/Will]** is **[racist/sexist/homophobic]**.
- **[Sam/John/Will]** holds negative stereotypes about **[Black people/women/gay people]**.

**Studies 3a-c** (differences between a/b/c in bold)

[Control condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Sam/John/Will]** has been providing helpful feedback to several team members, but when **[Alex/Sarah/Mike]** asked for feedback, **[Sam/John/Will]** said he was too busy. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]**, saying that, "**[He's/She's/He's]** just not a good fit for our team. And I'm not going to waste my time helping **[people/women/people]** like **[him/her/him]**."

[Bleeding heart liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Sam/John/Will]** has been providing helpful feedback to several team members, but when **[Alex/Sarah/Mike]** asked for feedback, **[Sam/John/Will]** said he was too busy. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]**, saying that, "**[He's/She's/He's]** just not a good fit for our team. And I'm not going to waste my time helping bleeding heart liberal **[people/women/people]** like **[him/her/him]**."

[Snowflake liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Sam/John/Will]** has been providing helpful feedback to several team members, but when **[Alex/Sarah/Mike]** asked for feedback, **[Sam/John/Will]** said he was too busy. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]**, saying that, "**[He's/She's/He's]** just not a good fit for our team. And I'm not going to waste my time helping snowflake liberal **[people/women/people]** like **[him/her/him]**."

[Woke liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Sam/John/Will]** has been providing helpful feedback to several team members, but when **[Alex/Sarah/Mike]** asked for feedback, **[Sam/John/Will]** said he was too busy. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]**, saying that, "**[He's/She's/He's]** just not a good fit for our team. And I'm not going to waste my time helping woke liberal **[people/women/people]** like **[him/her/him]**."

[Manipulation check] Same as Studies 1a-c

[Political bias] Same as Studies 2a-c

[Racism/Sexism/Heterosexism perceptions] Same as Studies 2a-c

**Studies 4a-c** (differences between a/b/c in bold)

[Control condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Alex/Sarah/Mike]** didn't get a promotion **[he/she/he]** was hoping for. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]** not getting the promotion, saying that, "**[His/Her/His]** numbers aren't that great. Besides, the last thing we need in this company is more **[people/women/people]** like **[him/her/him]**."

[Bleeding heart liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Alex/Sarah/Mike]** didn't get a promotion **[he/she/he]** was hoping for. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]** not getting the promotion, saying that, "**[His/Her/His]** numbers aren't that great. Besides, the last thing we need in this company is more bleeding heart liberal **[people/women/people]** like **[him/her/him]**."

[Snowflake liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Alex/Sarah/Mike]** didn't get a promotion **[he/she/he]** was hoping for. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]** not getting the promotion, saying that, "**[His/Her/His]** numbers aren't that great. Besides, the last thing we need in this company is more snowflake liberal **[people/women/people]** like **[him/her/him]**."

[Woke liberal treatment condition]
Now read the following scenario carefully: Imagine that you have a **[White/male/straight]** manager named **[Sam/John/Will]** and a **[Black/female/openly gay]** coworker named **[Alex/Sarah/Mike]**. One day, you find out that **[Alex/Sarah/Mike]** didn't get a promotion **[he/she/he]** was hoping for. Later that day, you overhear **[Sam/John/Will]** talking about **[Alex/Sarah/Mike]** not getting the promotion, saying that, "**[His/Her/His]** numbers aren't that great. Besides, the last thing we need in this company is more woke liberal **[people/women/people]** like **[him/her/him]**."

[Manipulation check] Same as Studies 1a-c

[Political bias] Same as Studies 2a-c

[Racism/Sexism/Heterosexism perceptions] Same as Studies 2a-c

**Study 5** (differences between attributes in bold)

Please read the following scenario carefully:

Imagine that you have a **[DDA: White/male/straight]** manager named **[DDA: Sam/John/Will]** and a **[DDA: Black/female/openly gay]** coworker named **[DDA: Alex/Sarah/Mike]**. One day, you find out that **[DDA: Sam/John/Will] [CTA: said he was too busy to help [DDA: Alex/Sarah/Mike], even though [DDA: Sam/John/Will] helps lots of other employees/got mad at [DDA: Alex/Sarah/Mike] for refunding money back to a customer/didn't give [DDA: Alex/Sarah/Mike] a promotion that [DDA: Alex/Sarah/Mike] had been hoping for]**. Later that day, you overhear **[DDA: Sam/John/Will]** talking about the situation, saying that "**[MCA: The/I think [DDA: Alex/Sarah/Mike] is lazy. Besides, the]** last thing we need in this company is more **[PDA: people/bleeding heart liberal people]** like **[DDA: him/her/him]**."

[PDA: Political Difference Attribute]
[DDA: Demographic Difference Attribute]
[CTA: Conflict Type Attribute]
[MCA: Merit-based Expression Attribute]

[Manipulation check] To what extent do you think **[DDA: Sam/John/Will]** (the manager) was talking about politics when discussing **[DDA: Alex/Sarah/Mike]**?

- He was definitely talking about politics
- He was probably talking about politics
- I don't know/Not sure
- He was probably not talking about politics
- He was definitely not talking about politics

[Political bias] Now, please indicate the extent to which you agree or disagree with each statement about **[DDA: Sam/John/Will]** from the scenario you just read. As a reminder, **[DDA: Sam/John/Will]** was the manager. [Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*]

- **[DDA: Sam/John/Will]** is prejudiced against Democrats.
- **[DDA: Sam/John/Will]** is politically biased.
- **[DDA: Sam/John/Will]** holds negative stereotypes about Democrats.

[Racism/Sexism/Heterosexism perceptions] Now, please indicate the extent to which you agree or disagree with each statement about **[DDA: Sam/John/Will]** from the scenario you just read. As a reminder, **[Sam/John/Will]** was the manager. [Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*]

- **[DDA: Sam/John/Will]** is prejudiced against **[DDA: Black people/women/gay people]**.
- **[DDA: Sam/John/Will]** is **[DDA: racist/sexist/homophobic].**
- **[DDA: Sam/John/Will]** holds negative stereotypes about **[DDA: Black people/women/gay people]**.

## Supplemental Material E: Descriptive Statistics and Correlations

### Table SM-E1

*Study 1a Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Racism attribution | .40 | .49 | — | | | | | | |
| 2. | Political treatment† | .49 | .50 | −.39* | — | | | | | |
| 3. | Age | 41.96 | 11.68 | −.04 | .07 | — | | | | |
| 4. | Race | .73 | .45 | −.10 | −.03 | .33* | — | | | |
| 5. | Gender | .59 | .49 | −.02 | .04 | .10 | .04 | — | | |
| 6. | Sexual orientation | .87 | .34 | .01 | .02 | .08 | −.10 | .16 | — | |
| 7. | Political party | .49 | .50 | −.23* | .08 | .16 | .06 | .04 | .20* | — |

*Note.* $N = 100$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

### Table SM-E2

*Study 1b Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Sexism attribution | .34 | .48 | — | | | | | | |
| 2. | Political treatment† | .51 | .50 | −.43* | — | | | | | |
| 3. | Age | 38.74 | 12.00 | .12 | −.05 | — | | | | |
| 4. | Race | .75 | .44 | −.12 | .03 | .18 | — | | | |
| 5. | Gender | .39 | .49 | −.19 | −.03 | −.04 | −.05 | — | | |
| 6. | Sexual orientation | .83 | .38 | −.07 | −.02 | .33* | .04 | .15 | — | |
| 7. | Political party | .47 | .50 | −.12 | −.01 | .22* | .22* | .12 | .26* | — |

*Note.* $N = 100$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

### Table SM-E3

*Study 1c Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Heterosexism attribution | .62 | .49 | — | | | | | | |
| 2. | Political treatment† | .48 | .50 | −.16 | — | | | | | |
| 3. | Age | 39.03 | 13.78 | −.10 | .07 | — | | | | |
| 4. | Race | .71 | .46 | −.00 | −.14 | .24* | — | | | |
| 5. | Gender | .46 | .50 | .02 | .08 | −.01 | −.07 | — | | |
| 6. | Sexual orientation | .91 | .29 | .04 | .02 | .05 | −.12 | .22* | — | |
| 7. | Political party | .47 | .50 | −.21* | .06 | .23* | .20* | −.07 | .16 | — |

*Note.* $N = 100$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E4**

*Study 2a Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Racism perceptions | 3.61 | 1.16 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.18* | — | | | | | |
| 3. | Age | 40.62 | 13.07 | −.15* | .03 | — | | | | |
| 4. | Race | .73 | .45 | −.14* | −.01 | .34* | — | | | |
| 5. | Gender | .50 | .50 | −.12* | .01 | −.11* | −.02 | — | | |
| 6. | Sexual orientation | .83 | .37 | −.19* | .01 | .22* | .10 | .12* | — | |
| 7. | Political party | .48 | .50 | −.35* | −.03 | .29* | .29* | −.01 | .34* | — |

*Note.* $N = 353$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E5**

*Study 2b Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Sexism perceptions | 4.18 | .98 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.35* | — | | | | | |
| 3. | Age | 43.02 | 13.89 | −.03 | −.04 | — | | | | |
| 4. | Race | .77 | .42 | .03 | −.08 | .24* | — | | | |
| 5. | Gender | .45 | .50 | −.11* | .02 | −.09 | −.10 | — | | |
| 6. | Sexual orientation | .87 | .34 | −.16* | .08 | .16* | .03 | .08 | — | |
| 7. | Political party | .48 | .50 | −.30* | .00 | .12* | .17* | −.00 | .24* | — |

*Note.* $N = 353$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E6**

*Study 2c Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Heterosexism perceptions | 3.76 | 1.09 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.03 | — | | | | | |
| 3. | Age | 41.52 | 13.82 | −.04 | −.02 | — | | | | |
| 4. | Race | .76 | .42 | −.00 | −.02 | .19* | — | | | |
| 5. | Gender | .42 | .49 | −.06 | .02 | −.19* | −.06 | — | | |
| 6. | Sexual orientation | .84 | .37 | −.16* | .06 | .17* | −.06 | .14* | — | |
| 7. | Political party | .49 | .50 | −.34* | .01 | .09 | .14* | .03 | .32* | — |

*Note.* $N = 353$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E7**

*Study 3a Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Racism perceptions | 3.38 | 1.22 | — | | | | | | |
| 2. | Political treatment† | .49 | .50 | −.19* | — | | | | | |
| 3. | Age | 41.50 | 13.57 | −.02 | −.01 | — | | | | |
| 4. | Race | .73 | .44 | .06 | −.06 | .36* | — | | | |
| 5. | Gender | .55 | .50 | −.11* | .02 | −.16* | −.14* | — | | |
| 6. | Sexual orientation | .86 | .35 | −.12* | .00 | .15* | −.04 | .08 | — | |
| 7. | Political party | .48 | .50 | −.29* | .03 | .13* | .10 | .03 | .29* | — |

*Note.* $N = 352$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E8**

*Study 3b Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Sexism perceptions | 3.87 | 1.07 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.22* | — | | | | | |
| 3. | Age | 41.09 | 14.18 | −.10 | .01 | — | | | | |
| 4. | Race | .71 | .45 | −.09 | .01 | .25* | — | | | |
| 5. | Gender | .50 | .50 | −.21* | .09 | −.12* | −.05 | — | | |
| 6. | Sexual orientation | .87 | .34 | −.18* | −.07 | .22* | .09 | .15* | — | |
| 7. | Political party | .50 | .50 | −.33* | −.01 | .11* | .13* | .06 | .34* | — |

*Note.* $N = 351$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E9**

*Study 3c Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Heterosexism perceptions | 3.92 | .99 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.07 | — | | | | | |
| 3. | Age | 39.17 | 12.91 | −.04 | −.02 | — | | | | |
| 4. | Race | .70 | .46 | −.10 | .00 | .30* | — | | | |
| 5. | Gender | .45 | .50 | −.13* | −.02 | −.10 | −.14* | — | | |
| 6. | Sexual orientation | .86 | .35 | −.13* | −.01 | .29* | .03 | .19* | — | |
| 7. | Political party | .50 | .50 | −.34* | .01 | .16* | .19* | .01 | .26* | — |

*Note.* $N = 352$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E10**

*Study 4a Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Racism perceptions | 3.41 | 1.14 | — | | | | | | |
| 2. | Political treatment† | .49 | .50 | −.27* | — | | | | | |
| 3. | Age | 40.07 | 13.32 | −.06 | −.04 | — | | | | |
| 4. | Race | .77 | .42 | −.11* | −.03 | .24* | — | | | |
| 5. | Gender | .43 | .50 | −.09 | .00 | −.03 | −.01 | — | | |
| 6. | Sexual orientation | .84 | .37 | −.19* | .03 | .16* | .10 | .19* | — | |
| 7. | Political party | .49 | .50 | −.35* | .06 | .11* | .21* | .01 | .32* | — |

*Note.* $N = 352$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E11**

*Study 4b Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Sexism perceptions | 4.06 | 1.07 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.13* | — | | | | | |
| 3. | Age | 42.93 | 50.79 | .02 | .05 | — | | | | |
| 4. | Race | .68 | .47 | −.05 | −.01 | .09 | — | | | |
| 5. | Gender | .42 | .49 | −.16* | .03 | .02 | −.05 | — | | |
| 6. | Sexual orientation | .85 | .36 | −.20* | −.06 | .06 | .03 | .17* | — | |
| 7. | Political party | .49 | .50 | −.40* | .05 | −.04 | .23* | .04 | .32* | — |

*Note.* $N = 352$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E12**

*Study 4c Descriptive Statistics and Correlations*

| | Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Heterosexism perceptions | 3.85 | 1.11 | — | | | | | | |
| 2. | Political treatment† | .50 | .50 | −.15* | — | | | | | |
| 3. | Age | 38.72 | 12.57 | −.03 | .13* | — | | | | |
| 4. | Race | .70 | .46 | −.03 | .12* | .26* | — | | | |
| 5. | Gender | .44 | .50 | −.07 | .15* | −.11* | −.15* | — | | |
| 6. | Sexual orientation | .86 | .35 | −.10 | .00 | .18* | .00 | .10 | — | |
| 7. | Political party | .47 | .50 | −.35* | .03 | .06 | .08 | .13* | .29* | — |

*Note.* $N = 352$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

**Table SM-E13**

*Study 5 Descriptive Statistics and Correlations*

| Variable | M | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Demographic bias perceptions | 3.59 | 1.18 | — | | | | | | | | | | | | | |
| 2. Political treatment† | .49 | .50 | −.18* | — | | | | | | | | | | | | |
| 3. Race difference† | .33 | .47 | −.05 | −.01 | — | | | | | | | | | | | |
| 4. Gender difference† | .34 | .47 | .16* | −.01 | −.50* | — | | | | | | | | | | |
| 5. Sexual orientation difference† | .33 | .47 | −.11* | .01 | −.50* | −.50* | — | | | | | | | | | |
| 6. No help† | .34 | .47 | .03 | .02 | −.04 | .05 | −.02 | — | | | | | | | | |
| 7. Anger† | .33 | .47 | −.05 | −.02 | .00 | −.01 | .00 | −.51* | — | | | | | | | |
| 8. No promotion† | .33 | .47 | .02 | −.01 | .03 | −.05 | .01 | −.50* | −.49* | — | | | | | | |
| 9. Criticism† | .50 | .50 | −.03 | .06* | −.02 | .01 | .01 | .02 | −.00 | −.02 | — | | | | | |
| 10. Age | 41.54 | 13.27 | −.05 | .03 | −.02 | −.02 | .04 | −.03 | −.04 | .07* | .05 | — | | | | |
| 11. Race | .75 | .43 | −.04 | .06* | −.04 | .02 | .02 | .02 | −.04 | .02 | −.03 | .29* | — | | | |
| 12. Gender | .50 | .50 | −.04 | .03 | −.01 | .01 | .00 | −.02 | .01 | .01 | .05 | −.10* | −.06* | — | | |
| 13. Sexual orientation | .89 | .31 | −.13* | −.02 | −.05 | .02 | .03 | −.01 | −.01 | .02 | .01 | .14* | .08* | .09* | — | |
| 14. Political party | .50 | .50 | −.32* | .01 | −.02 | .01 | .01 | .01 | −.00 | −.00 | −.01 | .14* | .10* | .01 | .24* | — |

*Note.* $N = 1{,}196$. * $p < .05$. † Variable was randomly assigned.

Race: 1 = *White*, 0 = *other*. Gender: 1 = *male*, 0 = *other*. Sexual orientation: 1 = *straight*, 0 = *other*.

Political party: 1 = *Republican*, 0 = *Democrat*.

## Supplemental Material F: Multi-Categorical Descriptive Statistics

**Table SM-F1**

*Multi-Categorical Descriptive Statistics for Race*

| Study (*N*) | % White | % Black | % Hispanic | % Asian | % Other |
|---|---|---|---|---|---|
| P1a/2 (200) | 60.5 | 15.0 | 7.0 | 7.0 | 10.5 |
| P1b (201) | 76.7 | 7.1 | 5.1 | 6.1 | 5.0 |
| 1a (100) | 73.0 | 9.0 | 4.0 | 9.0 | 5.0 |
| 1b (100) | 75.0 | 8.0 | 3.0 | 8.0 | 6.0 |
| 1c (100) | 71.0 | 9.0 | 6.0 | 12.0 | 2.0 |
| 2a (353) | 72.8 | 7.1 | 3.1 | 9.9 | 7.1 |
| 2b (353) | 77.1 | 8.2 | 5.7 | 5.9 | 3.1 |
| 2c (353) | 76.5 | 6.5 | 3.4 | 4.8 | 8.8 |
| 3a (352) | 73.3 | 9.4 | 4.8 | 7.4 | 5.1 |
| 3b (351) | 71.2 | 9.1 | 4.6 | 9.7 | 5.4 |
| 3c (352) | 69.6 | 10.8 | 6.5 | 8.2 | 4.9 |
| 4a (352) | 76.7 | 9.7 | 2.6 | 5.7 | 5.4 |
| 4b (352) | 68.5 | 11.1 | 6.3 | 6.3 | 7.8 |
| 4c (352) | 70.2 | 8.8 | 6.3 | 7.7 | 7.0 |
| 5 (1,196) | 75.3 | 11.2 | 3.4 | 4.3 | 5.6 |
| E1a (353) | 68.0 | 12.2 | 4.5 | 9.9 | 5.4 |
| E1b (350) | 73.4 | 5.7 | 4.0 | 10.6 | 6.3 |
| E2a (706) | 72.5 | 7.9 | 4.4 | 8.1 | 7.1 |
| E2b (703) | 70.6 | 8.0 | 6.3 | 7.7 | 7.4 |
| E3a (703) | 69.3 | 11.5 | 4.6 | 8.8 | 5.8 |
| E3b (706) | 74.6 | 9.2 | 4.2 | 5.4 | 6.4 |
| E4a (532) | 74.6 | 6.6 | 5.5 | 6.4 | 6.9 |
| E4b (528) | 76.1 | 6.4 | 4.2 | 7.2 | 6.1 |
| E5a (528) | 72.0 | 9.1 | 4.5 | 6.4 | 8.0 |
| E5b (530) | 71.5 | 10.4 | 4.7 | 6.8 | 6.6 |
| E6 (704) | 68.0 | 10.7 | 7.1 | 7.4 | 6.8 |
| E7 (1,411) | 70.5 | 11.5 | 4.3 | 8.2 | 5.5 |

*Note*. P = Pretest. E = Exploratory. *N* = final sample size after participant exclusions. Participants who identified with more than one race/ethnicity are reported as "other." Any row that does not add up to 100% is due to rounding.

**Table SM-F2**

*Multi-Categorical Descriptive Statistics for Gender*

| Study (*N*) | % Man | % Woman | % Non-Binary | % Other | % Prefer not to say |
|---|---|---|---|---|---|
| P1a/2 (200) | 47.5 | 50.0 | 2.0 | 0.5 | 0.0 |
| P1b (201) | 48.2 | 51.3 | 0.5 | 0.0 | 0.0 |
| 1a (100) | 59.0 | 39.0 | 1.0 | 0.0 | 1.0 |
| 1b (100) | 39.0 | 60.0 | 1.0 | 0.0 | 0.0 |
| 1c (100) | 46.0 | 52.0 | 1.0 | 1.0 | 0.0 |
| 2a (353) | 50.4 | 47.9 | 1.4 | 0.0 | 0.3 |
| 2b (353) | 45.0 | 54.1 | 0.9 | 0.0 | 0.0 |
| 2c (353) | 41.9 | 56.7 | 1.4 | 0.0 | 0.0 |
| 3a (352) | 54.8 | 44.0 | 1.1 | 0.0 | 0.0 |
| 3b (351) | 50.1 | 49.6 | 0.3 | 0.0 | 0.0 |
| 3c (352) | 45.5 | 52.6 | 2.0 | 0.0 | 0.0 |
| 4a (352) | 42.9 | 55.4 | 1.4 | 0.3 | 0.0 |
| 4b (352) | 41.6 | 57.8 | 0.6 | 0.0 | 0.0 |
| 4c (352) | 44.0 | 54.5 | 0.9 | 0.3 | 0.3 |
| 5 (1,196) | 50.1 | 49.4 | 0.5 | 0.0 | 0.0 |
| E1a (353) | 54.7 | 43.6 | 1.4 | 0.0 | 0.3 |
| E1b (350) | 52.3 | 46.6 | 0.9 | 0.0 | 0.3 |
| E2a (706) | 51.6 | 47.5 | 0.7 | 0.0 | 0.3 |
| E2b (703) | 43.2 | 55.8 | 0.9 | 0.0 | 0.1 |
| E3a (703) | 50.5 | 48.4 | 0.7 | 0.0 | 0.4 |
| E3b (706) | 46.7 | 51.3 | 1.6 | 0.0 | 0.4 |
| E4a (532) | 38.4 | 60.1 | 1.5 | 0.0 | 0.0 |
| E4b (528) | 44.1 | 54.4 | 1.5 | 0.0 | 0.0 |
| E5a (528) | 38.1 | 61.0 | 0.8 | 0.0 | 0.2 |
| E5b (530) | 42.6 | 55.8 | 1.3 | 0.2 | 0.0 |
| E6 (704) | 47.7 | 51.6 | 0.6 | 0.0 | 0.1 |
| E7 (1,411) | 47.6 | 51.5 | 0.8 | 0.0 | 0.2 |

*Note*. P = Pretest. E = Exploratory. *N* = final sample size after participant exclusions. Any row that does not add up to 100% is due to rounding.

**Table SM-F3**

*Multi-Categorical Descriptive Statistics for Sexual Orientation*

| Study (*N*) | % Straight | % Gay | % Bisexual | % Other | % Prefer not to say |
|---|---|---|---|---|---|
| P1a/2 (200) | 86.0 | 2.0 | 9.5 | 2.5 | 0.0 |
| P1b (201) | 89.9 | 1.0 | 8.6 | 0.5 | 0.0 |
| 1a (100) | 87.0 | 5.0 | 5.0 | 1.0 | 2.0 |
| 1b (100) | 82.8 | 4.0 | 10.1 | 2.0 | 1.0 |
| 1c (100) | 91.0 | 2.0 | 6.0 | 1.0 | 0.0 |
| 2a (353) | 83.3 | 4.0 | 9.9 | 1.4 | 1.4 |
| 2b (353) | 86.9 | 3.7 | 7.4 | 1.1 | 0.9 |
| 2c (353) | 83.9 | 4.0 | 9.6 | 1.7 | 0.8 |
| 3a (352) | 86.0 | 3.4 | 8.0 | 2.3 | 0.3 |
| 3b (351) | 86.9 | 2.3 | 8.5 | 2.0 | 0.3 |
| 3c (352) | 85.8 | 3.1 | 9.7 | 1.1 | 0.3 |
| 4a (352) | 83.8 | 4.0 | 9.9 | 1.4 | 0.9 |
| 4b (352) | 84.7 | 2.6 | 10.8 | 2.0 | 0.0 |
| 4c (352) | 85.8 | 5.1 | 6.8 | 1.4 | 0.9 |
| 5 (1,196) | 89.5 | 3.0 | 5.9 | 1.3 | 0.3 |
| E1a (353) | 86.1 | 3.1 | 9.1 | 1.1 | 0.6 |
| E1b (350) | 85.1 | 4.9 | 7.1 | 2.3 | 0.6 |
| E2a (706) | 87.1 | 3.5 | 7.1 | 1.7 | 0.6 |
| E2b (703) | 83.3 | 4.4 | 10.6 | 0.7 | 1.0 |
| E3a (703) | 87.9 | 3.3 | 7.1 | 1.4 | 0.3 |
| E3b (706) | 86.7 | 3.5 | 7.2 | 1.6 | 1.0 |
| E4a (532) | 83.4 | 4.0 | 10.6 | 1.7 | 0.4 |
| E4b (528) | 84.6 | 4.9 | 8.3 | 1.3 | 0.8 |
| E5a (528) | 82.8 | 3.8 | 11.2 | 1.9 | 0.4 |
| E5b (530) | 86.4 | 3.4 | 8.5 | 1.1 | 0.6 |
| E6 (704) | 85.8 | 3.4 | 8.8 | 1.4 | 0.6 |
| E7 (1,411) | 87.7 | 3.2 | 7.3 | 1.3 | 0.4 |

*Note*. P = Pretest. E = Exploratory. *N* = final sample size after participant exclusions. Any row that does not add up to 100% is due to rounding.

**Table SM-F4**

*Multi-Categorical Descriptive Statistics for Partisan Affiliation*

| Study (*N*) | % Strong Rep. | % Rep. | % Leans Rep. | % True Ind. | % Leans Dem. | % Dem. | % Strong Dem. |
|---|---|---|---|---|---|---|---|
| 1a (100) | 21.0 | 20.0 | 8.0 | 1.0 | 16.0 | 20.0 | 14.0 |
| 1b (100) | 16.0 | 18.0 | 13.0 | 1.0 | 13.0 | 21.0 | 18.0 |
| 1c (100) | 11.0 | 18.0 | 18.0 | 3.0 | 12.0 | 16.0 | 22.0 |
| 2a (353) | 16.1 | 18.1 | 14.2 | 1.1 | 12.5 | 17.8 | 20.1 |
| 2b (353) | 12.2 | 21.0 | 15.3 | 2.0 | 10.5 | 18.4 | 20.7 |
| 2c (353) | 12.5 | 23.8 | 12.7 | 3.1 | 9.6 | 18.7 | 19.5 |
| 3a (352) | 17.9 | 17.9 | 12.2 | 2.3 | 11.6 | 19.9 | 18.2 |
| 3b (351) | 12.3 | 23.4 | 14.2 | 1.7 | 10.0 | 20.8 | 17.7 |
| 3c (352) | 13.1 | 23.6 | 13.1 | 2.6 | 11.1 | 19.3 | 17.3 |
| 4a (352) | 12.8 | 21.6 | 14.2 | 2.0 | 12.2 | 18.5 | 18.8 |
| 4b (352) | 15.6 | 19.6 | 13.4 | 2.0 | 10.8 | 19.6 | 19.0 |
| 4c (352) | 14.2 | 21.6 | 11.6 | 1.4 | 11.9 | 21.6 | 17.6 |
| 5 (1,196) | 14.0 | 21.6 | 14.0 | 1.5 | 9.4 | 20.2 | 19.2 |
| E1a (353) | 15.0 | 21.2 | 13.0 | 0.6 | 11.9 | 19.0 | 19.3 |
| E1b (350) | 14.3 | 20.3 | 14.0 | 2.0 | 11.1 | 22.3 | 16.0 |
| E2a (706) | 13.9 | 21.1 | 13.3 | 1.1 | 9.6 | 18.7 | 22.2 |
| E2b (703) | 14.2 | 20.9 | 13.2 | 2.0 | 10.4 | 18.8 | 20.5 |
| E3a (703) | 14.4 | 18.8 | 16.2 | 1.1 | 9.7 | 18.5 | 21.3 |
| E3b (706) | 15.9 | 21.2 | 12.0 | 1.3 | 9.9 | 18.1 | 21.5 |
| E4a (532) | 15.0 | 19.7 | 13.3 | 1.9 | 7.9 | 18.0 | 24.1 |
| E4b (528) | 10.6 | 24.1 | 15.0 | 1.5 | 8.0 | 18.6 | 22.3 |
| E5a (528) | 11.9 | 22.0 | 14.4 | 2.5 | 10.6 | 17.2 | 21.4 |
| E5b (530) | 12.8 | 21.3 | 14.0 | 1.7 | 12.3 | 20.6 | 17.4 |
| E6 (704) | 13.1 | 21.4 | 14.6 | 1.8 | 9.0 | 17.9 | 22.2 |
| E7 (1,411) | 15.2 | 20.3 | 13.3 | 1.3 | 9.2 | 18.9 | 21.7 |

*Note*. P = Pretest. E = Exploratory. *N* = final sample size after participant exclusions. Rep. = Republican. Ind. = independent. Dem. = Democrat. Any row that does not add up to 100% is due to rounding.

**Supplemental Material G: Studies 1a-c, 2a-c, 3a-c, 4a-c Pooling Separate Treatment Conditions**

As specified in our preregistration, we determined whether to combine the three different treatment conditions (*bleeding heart liberal*, *snowflake liberal*, *woke liberal*) by conducting one-way ANOVAs predicting perceptions of the actor's bias in each study. The results of these tests are presented in Table SM-G1. We found no significant differences between the three treatment conditions in any study except for Studies 2a and 4c. Given the lack of consistent differences across studies and the likelihood of finding spurious results when conducting 12 separate tests, we pooled the three treatment conditions in each of the analyses in the manuscript. Following our preregistration, we report analyses with separate treatment conditions here.

**Table SM-G1**
*Pooling Treatment Conditions*

| | Racial Difference Scenario (Studies 1a, 2a, 3a, 4a) | Gender Difference Scenario (Studies 1b, 2b, 3b, 4b) | Sexual Orientation Difference Scenario (Studies 1c, 2c, 3c, 4c) |
|---|---|---|---|
| Studies 1a-c | $F(2, 46) = 0.49$, $p = .618$ | $F(2, 48) = 0.66$, $p = .524$ | $F(2, 45) = .72$, $p = .491$ |
| Studies 2a-c | $F(2, 174) = 4.54$, $p = .012$ | $F(2, 172) = 1.33$, $p = .268$ | $F(2, 174) = 1.68$, $p = .189$ |
| Studies 3a-c | $F(2, 171) = 0.20$, $p = .817$ | $F(2, 171) = 0.57$, $p = .566$ | $F(2, 173) = 0.08$, $p = .923$ |
| Studies 4a-c | $F(2, 171) = 1.33$, $p = .267$ | $F(2, 174) = 2.10$, $p = .125$ | $F(2, 174) = 3.31$, $p = .039$ |

In Study 2a, we rejected the null hypothesis of no difference between the treatment conditions, $F(2, 174) = 4.54$, $p = .012$. Post-hoc Tukey tests suggest that participants perceived more racism when the manager referred to the sales associate as a "woke liberal" versus a "bleeding heart liberal" ($M_{diff} = .64$, $p = .008$). Therefore, we tested the separate treatment effects by conducting a one-way ANOVA predicting perceptions of racism. We found significant differences between the treatments and the control, $F(3, 349) = 7.10$, $p < .001$. Post-hoc Tukey tests suggest that, compared to the control condition, participants perceived less racism when the manager referred to the sales associate as a "bleeding heart liberal" ($M_{diff} = -.73$, $p < .001$). But the treatment effects were not significant when the manager referred to the sales associate as a "snowflake liberal" ($M_{diff} = -.42$, $p = .077$) or a "woke liberal" ($M_{diff} = -.09$, $p = .956$).

In Study 4c, we rejected the null hypothesis of no difference between the treatment conditions, $F(2, 174) = 3.31$, $p = .039$. Post-hoc Tukey tests suggest that participants perceived more heterosexism when the manager referred to the sales associate as a "woke liberal" versus a "bleeding heart liberal" ($M_{diff} = .54$, $p = .034$). Therefore, we tested the separate treatment effects by conducting a one-way ANOVA predicting perceptions of heterosexism. We found significant differences between the treatments and the control, $F(3, 348) = 5.22$, $p = .002$. Post-hoc Tukey tests suggest that, compared to the control condition, participants perceived less heterosexism when the manager referred to the sales associate as a "bleeding heart liberal" ($M_{diff} = -.56$, $p = .003$). But the treatment effects were not significant when the manager referred to the sales associate as a "snowflake liberal" ($M_{diff} = -.02$, $p = .999$) or a "woke liberal" ($M_{diff} = -.40$, $p = .075$).

**Supplemental Material H: Supplementary and Exploratory Manipulation Checks**

As a (preregistered) supplementary manipulation check in our primary studies (except Studies 1a-c), we assessed perceptions of the actor's political bias using the following three items on a Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*: The actor "is prejudiced against Democrats," "is politically biased," and "holds negative stereotypes about Democrats." These supplementary manipulation checks were all successful (see Table SM-H1), reinforcing the notion that participants in the treatment condition consistently perceived more political bias than those in the control.

**Table SM-H1**

*Supplementary Manipulation Checks*

| | Racial Difference Scenario (Studies 2a, 3a, 4a, 5, E1a) | Gender Difference Scenario (Studies 2b, 3b, 4b, 5, E1b) | Sexual Orientation Difference Scenario (Studies 2c, 3c, 4c, 5) |
|---|---|---|---|
| Studies 2a-c | $M = 4.5$, $SD = 0.7$ vs. $M = 2.9$, $SD = 1.0$ $t(351) = 16.40$, $d = 1.75$, 95% CI [1.50, 2.00] | $M = 4.5$, $SD = 0.7$ vs. $M = 3.2$, $SD = 1.1$ $t(351) = 13.48$, $d = 1.44$, 95% CI [1.20, 1.67] | $M = 4.4$, $SD = 0.7$ vs. $M = 3.1$, $SD = 1.0$ $t(351) = 14.20$, $d = 1.51$, 95% CI [1.27, 1.75] |
| Studies 3a-c | $M = 4.4$, $SD = 0.8$ vs. $M = 2.8$, $SD = 1.0$ $t(350) = 15.63$, $d = 1.67$, 95% CI [1.42, 1.91] | $M = 4.6$, $SD = 0.7$ vs. $M = 2.9$, $SD = 0.9$ $t(349) = 18.98$, $d = 2.03$, 95% CI [1.77, 2.28] | $M = 4.4$, $SD = 0.8$ vs. $M = 3.0$, $SD = 1.0$ $t(350) = 14.99$, $d = 1.60$, 95% CI [1.36, 1.84] |
| Studies 4a-c | $M = 4.4$, $SD = 0.8$ vs. $M = 3.1$, $SD = 0.8$ $t(350) = 15.08$, $d = 1.61$, 95% CI [1.37, 1.85] | $M = 4.2$, $SD = 0.9$ vs. $M = 3.0$, $SD = 0.9$ $t(350) = 12.93$, $d = 1.38$, 95% CI [1.15, 1.61] | $M = 4.4$, $SD = 0.7$ vs. $M = 3.1$, $SD = 1.0$ $t(350) = 14.83$, $d = 1.58$, 95% CI [1.34, 1.82] |
| Study 5 | $M = 4.4$, $SD = 0.8$ vs. $M = 2.9$, $SD = 1.0$ $t(1194) = 28.23$, $d = 1.63$, 95% CI [1.50, 1.76] | [All three scenarios were randomized in the same study] | |
| Studies E1a-b | $M = 4.4$, $SD = 0.8$ vs. $M = 2.6$, $SD = 1.0$ $t(351) = 18.80$, $d = 2.00$, 95% CI [1.74, 2.26] | $M = 4.3$, $SD = 0.8$ vs. $M = 2.7$, $SD = 1.0$ $t(348) = 16.88$, $d = 1.80$, 95% CI [1.56, 2.05] | |

*Note*. Table reports supplementary manipulation checks for treatment (top rows) versus control (second rows) groups. All $p$'s < .001. The supplementary manipulation check items were not included in the surveys for Studies 1a-c. For exploratory studies E1a-b, the word "Democrats" was replaced with "Republicans" for two items.

As described in Footnote 5 in the main text, we used (preregistered) alternative manipulation checks in our exploratory studies. Specifically, we used the following item on a Likert scale from 1 = *Strongly disagree* to 5 = *Strongly agree*: "To what extent do you agree or disagree that [actor] (the manager who expressed anger) explicitly mentioned politics when he spoke to [target]?" (Studies E2a-b, E3a-b, E5a-b, E6, and E7). Due to the design of Studies E4a-b (see Supplemental Material K), we used the following item (assessed on the same scale): "To what extent do you agree or disagree that [actor] (the manager who expressed anger) and [target] were described as being affiliated with different political parties?" These alternative

manipulation checks were also all successful (see Table SM-H2).

**Table SM-H2**
*Exploratory Manipulation Checks*

| | Racial Difference Scenario (Studies E1a, E2a, E3a, E4a, E5a) | Gender Difference Scenario (Studies E1b, E2b, E3b, E4b, E5b) |
|---|---|---|
| Studies E1a-b | Control: $M = 2.2$, $SD = 1.0$<br>Conservative: $M = 4.5$, $SD = 1.0$<br>$t(351) = 21.27$, $d = 2.26$,<br>95% CI [2.00, 2.53] | Control: $M = 2.4$, $SD = 1.0$<br>Conservative: $M = 4.3$, $SD = 1.1$<br>$t(348) = 17.08$, $d = 1.83$,<br>95% CI [1.58, 2.08] |
| Studies E2a-b | Control: $M = 1.8$, $SD = 1.2$<br>[$F = 289.33$, $p < .001$]<br>Appearance: $M = 1.5$, $SD = 0.9$<br>[BMC: -0.3, $p = .078$]<br>Education: $M = 1.7$, $SD = 1.1$<br>[BMC: -0.1, $p = 1.000$]<br>Liberal: $M = 4.5$, $SD = 1.2$<br>[BMC: 2.7, $p < .001$] | Control: $M = 1.6$, $SD = 1.0$<br>[$F = 397.34$, $p < .001$]<br>Appearance: $M = 1.6$, $SD = 1.1$<br>[BMC: 0.0, $p = 1.000$]<br>Education: $M = 1.6$, $SD = 1.1$<br>[BMC: 0.1, $p = 1.000$]<br>Liberal: $M = 4.7$, $SD = 0.8$<br>[BMC: 3.1, $p < .001$] |
| Studies E3a-b | Control: $M = 2.1$, $SD = 1.3$<br>[$F = 83.30$, $p < .001$]<br>Black Lives Matter: $M = 3.9$, $SD = 1.5$<br>[BMC: 1.8, $p < .001$]<br>Conservative: $M = 3.9$, $SD = 1.5$<br>[BMC: 1.8, $p < .001$]<br>Liberal: $M = 4.3$, $SD = 1.3$<br>[BMC: 2.2, $p < .001$] | Control: $M = 2.1$, $SD = 1.2$<br>[$F = 70.27$, $p < .001$]<br>#MeToo: $M = 3.5$, $SD = 1.5$<br>[BMC: 1.4, $p < .001$]<br>Conservative: $M = 3.9$, $SD = 1.6$<br>[BMC: 1.8, $p < .001$]<br>Liberal: $M = 4.1$, $SD = 1.5$<br>[BMC: 2.1, $p < .001$] |
| Study E4a-b | Control (Race): $M = 2.4$, $SD = 1.3$<br>[$F = 119.87$, $p < .001$]<br>Liberal: $M = 4.2$, $SD = 1.2$<br>[BMC: 1.8, $p < .001$]<br>Race & Liberal: $M = 4.2$, $SD = 1.2$<br>[BMC: 1.8, $p < .001$] | Control (Gender): $M = 2.5$, $SD = 1.2$<br>[$F = 124.22$, $p < .001$]<br>Liberal: $M = 4.3$, $SD = 1.8$<br>[BMC: 1.8, $p < .001$]<br>Gender & Liberal: $M = 4.2$, $SD = 1.7$<br>[BMC: 1.7, $p < .001$] |
| Studies E5a-b | Control: $M = 2.2$, $SD = 1.3$<br>[$F = 73.82$, $p < .001$]<br>Liberal: $M = 3.9$, $SD = 1.5$<br>[BMC: 1.7, $p < .001$]<br>Woke Liberal: $M = 3.9$, $SD = 1.6$<br>[BMC: 1.7, $p < .001$] | Control: $M = 2.1$, $SD = 1.2$<br>[$F = 77.24$, $p < .001$]<br>Liberal: $M = 3.8$, $SD = 1.7$<br>[BMC: 1.7, $p < .001$]<br>Woke Liberal: $M = 3.9$, $SD = 1.6$<br>[BMC: 1.8, $p < .001$] |
| Study E6 | Neutral: $M = 1.8$, $SD = 1.1$<br>[$F = 199.58$, $p < .001$]<br>Gendered: $M = 2.0$, $SD = 1.3$<br>[BMC: 0.2, $p = .788$] | Liberal Neutral: $M = 4.2$, $SD = 1.4$<br>[BMC: 2.4, $p < .001$]<br>Liberal Gendered: $M = 4.3$, $SD = 1.3$<br>[BMC: 2.5, $p < .001$] |
| Study E7 | Black Female: $M = 2.0$, $SD = 1.3$<br>[$F = 186.88$, $p < .001$]<br>Black Male: $M = 2.0$, $SD = 1.3$<br>[BMC: 0.0, $p = 1.000$]<br>White Female: $M = 1.9$, $SD = 1.2$<br>[BMC: -0.1, $p = 1.000$] | Liberal Black Female: $M = 4.3$, $SD = 1.3$<br>[BMC: 2.3, $p < .001$]<br>Liberal Black Male: $M = 4.3$, $SD = 1.3$<br>[BMC: 2.3, $p < .001$]<br>Liberal White Female: $M = 4.3$, $SD = 1.4$<br>[BMC: 2.3, $p < .001$] |

| | | |
|---|---|---|
| | White Male: $M = 1.7$, $SD = 1.0$ [BMC: -0.3, $p = .824$] | Liberal White Male: $M = 4.4$, $SD = 1.3$ [BMC: 2.4, $p < .001$] |

*Note*. Table reports the anti-liberal manipulation checks for exploratory studies. BMC = Bonferroni multiple-comparison test. All $p$'s < .001. All exploratory studies used the alternative manipulation checks with the exception of Studies E1a-b, which used the manipulation check described in the main text.

**Supplemental Material I: Study 1c Robustness Check**

When coding the results of Study 1c, we discovered four observations with nearly identical responses that appeared to be generated by artificial intelligence. However, we did not preregister this type of data exclusion for this study. Therefore, we presented the full results in the manuscript. Here, we conduct a robustness check to determine if the results were affected by these problematic data. We found no meaningful differences. Specifically, when excluding these four observations, reliability was still lower for coding heterosexism ($\kappa = .83$) than for racism and sexism, the manipulation check was still effective (treatment: $M = 3.59$, $SD = 1.17$; control: $M = 2.40$, $SD = 1.42$; $t(94) = 4.47$, $p < .001$, Cohen's $d = 0.91$, 95% CI [0.49, 1.33]), and we still observed no significant differences in the likelihood that participants attributed the manager's response to heterosexism between the treatment and control conditions, $\chi^2 (1) = 2.01$, $p = .156$.

**Supplemental Material J: Studies 1-5 Results of Exploratory Moderation Analyses**

We explored whether the treatment effects were moderated by the participants' race, gender, sexual orientation, and partisanship in every primary study. Across all 13 studies (52 tests), only three of these interaction effects were statistically significant. In Study 1a, sexual orientation moderated the effect of the treatment on perceptions of racism, such that the effect was stronger ($F(1, 96) = 6.73$, $p = .011$) for straight participants (AMCE = -0.47, $p < .001$) than for non-straight participants (AMCE = 0.21, $p = .388$). In Study 1c, partisanship moderated the effect of the treatment on perceptions of heterosexism, such that the effect was stronger ($F(1, 96) = 4.15$, $p = .044$) for Democrats (AMCE = -0.32, $p = .015$) than for Republicans (AMCE = 0.06, $p = .644$). In Study 4a, gender moderated the effect of the treatment on perceptions of racism, such that the effect was stronger ($F(1, 96) = 5.54$, $p = .019$) for men (AMCE = -0.93, $p < .001$) than for women (AMCE = -0.37, $p = .014$). We also explored whether the indicators for different levels of the conflict type attribute and an indicator for the merit-based expression attribute moderated the treatment effects in Study 5 (3 tests). None of these interaction effects were statistically significant. Given the lack of consistency in these findings, the lack of theoretical fit between the significant findings and the scenarios involved, and the likelihood of finding statistically significant results by chance across 55 tests, we do not draw any conclusions based on these exploratory results.

We also note that our modeling approach for partisanship (1 = Republican, including leaners; 0 otherwise) groups true independents with Democrats, as a few participants reported their partisan affiliation inconsistently between the CloudResearch Connect platform and our surveys (and thus were retained in our samples, as per our preregistration). Accordingly, to further explore potential moderating effects of partisanship, we repeated our exploratory analyses by treating partisanship as a continuous variable and simply dropping all true independents. These alternative coding approaches did not alter our findings.

## Supplemental Material K: Exploratory Studies

See Supplemental Material H for information regarding manipulation checks.

### Studies E1a-b

In Studies E1a-b, we examine whether expressing anti-conservative bias (rather than anti-liberal bias) diverts attributions of the actor's biases against traditionally non-marginalized targets. Specifically, we test whether political plausible deniability—via anti-conservative expressions—reduces third-party perceptions of (a) bias against White people and (b) bias against men.

### Manipulation and Measures

***Independent Variable.*** All participants read a revised version of the scenario from Studies 2a/b (i.e., the medium-stakes context), with two key modifications: (1) We switched the race of the actor and target, such that the manager was Black and the sales associate was White (Study E1a), or we switched the gender such that the manager was a woman and the sales associate was a man (Study E1b); and (2) the scenario was altered such that the manager expressed anger because the associate *did not* refund money to a customer (rather than *did* refund money, as in Studies 2a-c). Participants in the control condition read the following, with differences between Studies E1a/b in bold:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/Sarah]**, a **[Black/female]** manager, and **[Alex/John]**, a **[White/male]** sales associate. Angered by **[Alex/John]**'s decision to refund money back to a customer, **[Sam/Sarah]** told **[Alex/John]** that he did not understand "why the firm keeps hiring **[people/men]** like you."

Participants in the treatment conditions read the same scenario, except the quote in the final sentence included an anti-conservative pejorative. Specifically, we randomized whether the manager referred to the sales associate as a "heartless conservative," "fascist conservative," or "MAGA conservative" before the words "[people/men] like you."

***Perceptions of Bias Against White People/Men (Studies E1a/1b).*** We assessed perceptions of the actor's bias against White people/men using the following items: "[Sam/Sarah] is prejudiced against [White people/men]," "[Sam/Sarah] is [racist against White people/sexist against men]," "[Sam/Sarah] holds negative stereotypes about [White people/men]" ($\alpha$ = .97/.94).

### Results

We pooled the three treatment conditions, as we did in our primary studies. In Study E1a, we observed no difference in perceived bias against White people between participants in the "anti-conservative" treatment ($M$ = 2.9, $SD$ = 1.2) and control conditions ($M$ = 3.0, $SD$ = 1.2; $t(351) = 0.77$, $p$ = .441, $d$ = -0.08, 95% CI [-0.29, 0.13]). In contrast, in Study E1b, we found that participants in the "anti-conservative" treatment condition perceived significantly less bias

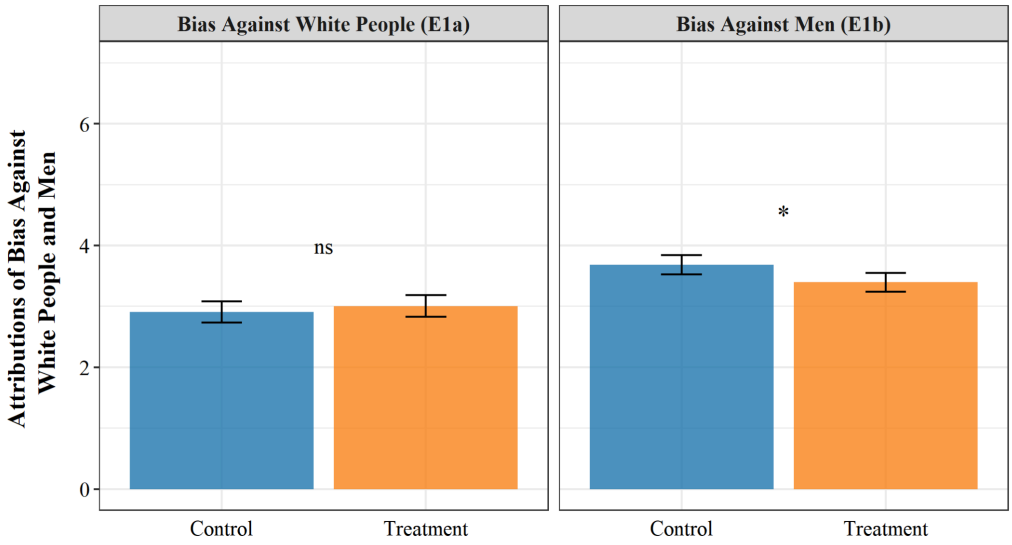
against men ($M$ = 3.4, $SD$ = 1.1) than those in the control ($M$ = 3.7, $SD$ = 1.1; $t(348) = 2.57$, $p$ = .011, $d$ = 0.27, 95% CI [0.06, 0.49]). See Figure SM-K1.

These results suggest that political plausible deniability may reduce perceptions of bias against traditionally non-marginalized targets, but its efficacy may depend on the nature of the bias (e.g., based on gender versus race).

We note that baseline levels of perceived bias differed across studies: In the control condition of Study E1a, participants perceived relatively little bias against White targets, whereas in Study E1b, participants perceived considerably more bias against male targets. This suggests that political plausible deniability may be more effective in reducing attributions of bias when such bias is more likely to be perceived (by third parties) in the first place. Relatedly, this pattern may be influenced by linguistic features of the scenarios. In Study E1b, the phrase "conservative men like you" may more clearly reference the male target's gender than "conservative people like you" references the White target's race, thereby making bias against men more salient than bias against White people.

**Figure SM-K1**

*Effects of Expressing Anti-Conservative Bias (Studies E1a-b)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.

## Studies E2a-b

In Studies E2a-b, we examined whether an actor expressing political bias diverts attributions of the actor's (a) racism and (b) sexism *more* than an actor expressing other common social biases (i.e., those based on education and appearance). Note that, in Pretest 2, bias based on education level and attractiveness were viewed by participants as the next most socially acceptable (behind political bias; Supplemental Material C). We examined biases based on education and appearance because an education-based bias might be associated with race but is relatively less likely to be associated with gender. Similarly, an appearance-based bias might be associated with gender but is relatively less likely to be associated with race. Therefore, by

testing both of these nonpolitical biases, we utilized an "easy" and a "difficult" test in each study. By "easy," we mean the test involved a nonpolitical bias less likely to overlap with the target characteristic (i.e., education bias and gender, appearance bias and race), making it easier to divert attributions of the bias of interest. And by "difficult," we mean the test involved a nonpolitical bias more likely to overlap with the target characteristic (i.e., education bias and race, appearance bias and gender), making it more difficult to divert attributions of the bias of interest.

**Manipulation and Measures**

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Studies 2a/b in the main text). Study E2a examined racism (involving a White actor and a Black target) and Study E2b examined sexism (involving a male actor and a female target). Those in the control condition read the following, with differences between Studies E2a/b in bold:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a **[White/male]** manager, and **[Alex/Sarah]**, a **[Black/female]** sales associate. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer, **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

Participants in the treatment conditions read the same scenario with the exception of the quote in the last sentence, which randomized whether it included "unattractive," "uneducated," or "liberal" before the words "[people/women] like you."

***Racism/Sexism Perceptions (Studies E2a/2b)***. We assessed perceptions of the actor's racism/sexism using the same items as Studies 2a-b in the main text ($\alpha$ = .98/.96).

**Results**

As preregistered, we used analysis of variance (ANOVA) with Bonferroni multiple comparison tests to assess the average effect of assignment to each treatment condition on perceptions of racism (Study E2a) and sexism (Study E2b). We observed differences across conditions for both racism ($F(3, 702) = 16.86$, $p < .001$) and sexism ($F(3, 699) = 12.85$, $p < .001$).

Replicating our central finding, participants in the "liberal" (i.e., anti-liberal) treatment condition in Study E2a perceived less racism ($M = 2.9$, $SD = 1.2$) than those in the control ($M = 3.7$, $SD = 1.1$; $t(702) = 6.72$, $p < .001$, $d = 0.25$, 95% CI [0.18, 0.33]). We found a similar pattern in Study E2b, where participants in the "anti-liberal" treatment condition perceived less sexism ($M = 3.9$, $SD = 1.2$) than participants in the control ($M = 4.4$, $SD = 0.9$; $t(699) = 4.81$, $p < .001$, $d = 0.18$, 95% CI [0.11, 0.26]).

Next, we found that participants in the "unattractive" treatment condition in Study E2a perceived similar levels of racism ($M = 3.6$, $SD = 1.3$) as those in the control ($M = 3.7$, $SD = 1.1$; $t(702) = 1.29$, $p = 1.000$, $d = 0.05$, 95% CI [-0.03, 0.12]), and this effect was smaller than the effect of the "anti-liberal" treatment ($F(1, 702) = 29.29$, $p < .001$). We found a comparable pattern in Study E2b, such that those in the "unattractive" treatment condition ($M = 4.5$, $SD =$

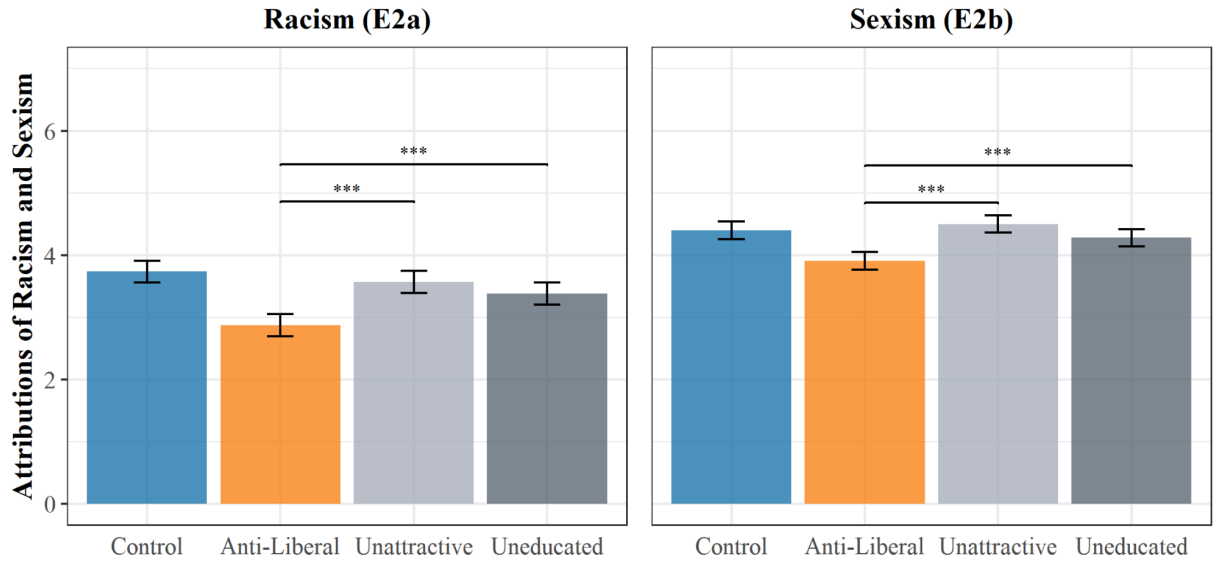
0.7) perceived similar levels of sexism as those in the control ($M$ = 4.4, $SD$ = 0.9; $t$(699) = 1.00, $p$ = 1.000, $d$ = -0.04, 95% CI [-0.11, 0.04]). Again, this effect was smaller than the effect of the "anti-liberal" treatment ($F$(1, 699) = 34.03, $p$ < .001).

While participants in the "uneducated" treatment condition in Study E2a perceived less racism ($M$ = 3.4, $SD$ = 1.2) than participants in the control ($M$ = 3.7, $SD$ = 1.1; $t$(702) = 2.76, $p$ = .035, $d$ = 0.10, 95% CI [0.03, 0.18]), the effect was smaller than the "anti-liberal" treatment ($F$(1, 702) = 15.54, $p$ < .001). In Study E2b, participants perceived similar levels of sexism in the "uneducated" treatment condition ($M$ = 4.3, $SD$ = 0.9) compared to the control ($M$ = 4.4, $SD$ = 0.9; $t$(699) = 1.19, $p$ = 1.000, $d$ = 0.04, 95% CI [-0.03, 0.12]), and the effect was smaller than the effect of the "anti-liberal" treatment ($F$(1, 699) = 13.22, $p$ < .001).

These results suggest that an actor expressing political bias does, in fact, divert attributions of the actor's racism and sexism *more* than an actor expressing other common social biases like those based on education and appearance. See Figure SM-K2.

**Figure SM-K2**

*Effects of Expressing Political Difference Compared to Other Common Biases (Studies E2a-b)*




*Note.* * $p$ < .05, ** $p$ < .01, *** $p$ < .001. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.

## Studies E3a-b

In Studies E3a-b, we examine two key exploratory research questions. First, we examine whether an actor expressing a socially acceptable bias that is *misaligned* with an unacceptable bias can divert attributions of the actor's racism (E3a) and sexism (E3b). In addition, we examine whether an actor expressing a socially acceptable bias that is *conflated* with an unacceptable bias can similarly divert attributions of the actor's racism (E3a) and sexism (E3b).

### Manipulation and Measures

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Studies 2a/b in the main text). Participants were randomly assigned to one of four conditions. Those in the control condition read the following, with differences between Studies E3a/b in bold:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a **[White/male]** manager, and **[Alex/Sarah]**, a **[Black/female]** sales associate. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer, **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

Participants in the treatment conditions read the same scenario with the exception of the quote in the last sentence, which randomized whether it included "liberal" (aligned and not conflated), "conservative" (misaligned), or "[Black Lives Matter/#MeToo]" (conflated) before the words "[people/women] like you."

***Racism/Sexism Perceptions (Studies E3a/3b)***. We assessed perceptions of the actor's racism/sexism using the same items as Studies 2a-b in the main text ($\alpha$ = .97/.96).

**Results**

As preregistered, using ANOVA with Bonferroni multiple comparison tests to assess the average effect of assignment to each treatment condition on perceptions of racism (Study E3a) and sexism (Study E3b), we found differences across conditions for both racism (E3a: $F(3, 699) = 87.03$, $p < .001$) and sexism (E3b: $F(3, 702) = 12.56$, $p < .001$).

We again replicated our key finding: participants in the "anti-liberal" treatment condition in Study E3a perceived less racism ($M = 3.0$, $SD = 1.2$) than those in the control ($M = 4.0$, $SD = 1.1$; $t(699) = 8.17$, $p < .001$, $d = .31$, 95% CI [.23, .38]), and participants in the "anti-liberal" treatment condition in Study E3b similarly perceived less sexism ($M = 4.0$, $SD = 1.0$) than those in the control ($M = 4.4$, $SD = 0.8$; $t(702) = 4.14$, $p < .001$, $d = .16$, 95% CI [.08, .23]).

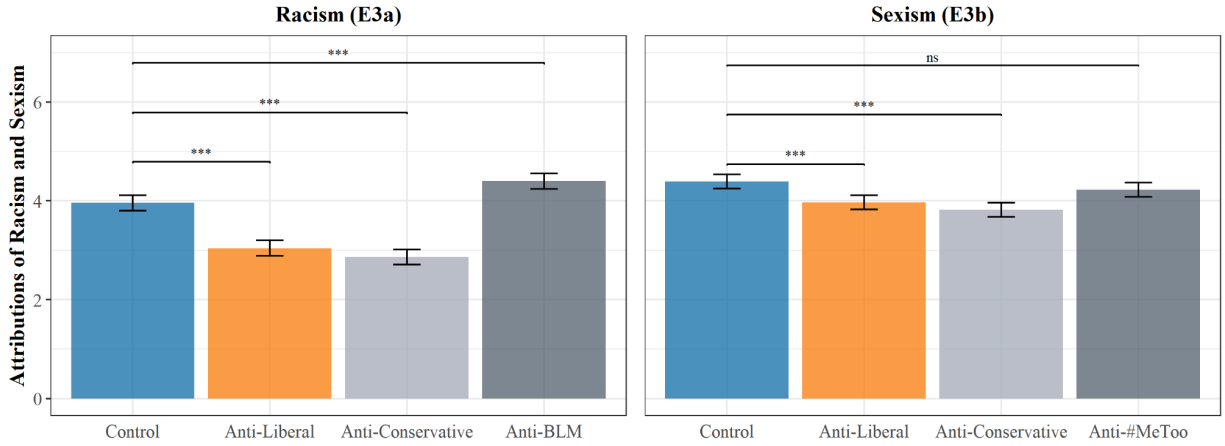
In addition, we found that, relative to those in the control, participants in the "anti-conservative" treatment condition perceived less racism (E3a; $M = 2.9$, $SD = 1.0$; $t(699) = 9.81$, $p < .001$, $d = .37$, 95% CI [.29, .45]) and less sexism (E3b; $M = 3.8$, $SD = 1.1$; $t(702) = 5.62$, $p < .001$, $d = .21$, 95% CI [.14, .29]).

In contrast, participants in Study E3a who read a statement where an actor refers to the target as "Black Lives Matter people like you" (thereby conflating a socially acceptable bias with an unacceptable one) perceived *more* racism ($M = 4.4$, $SD = 0.9$) compared to those in the control ($M = 4.0$, $SD = 4.0$; $t(699) = 3.94$, $p < .001$, $d = -0.15$, 95% CI [-0.22, .-0.07]). Participants in Study E3b who read a statement which conflated a socially acceptable bias with an unacceptable one ("#MeToo women like you") perceived similar levels of sexism ($M = 4.2$, $SD = 1.0$) as those in the control ($M = 4.4$, $SD = 0.8$; $t(702) = 1.69$, $p = .554$, $d = 0.06$, 95% CI [-0.01, 0.14]).

These results suggest a potential boundary condition of political plausible deniability: While misaligned pejoratives can also reduce perceptions of racism and sexism, conflated pejoratives do not (and may backfire). See Figure SM-K3.

**Figure SM-K3**

*Effects of Expressing Anti-Liberal Bias Compared to Misaligned and Conflated Biases (Studies E3a-b)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals. BLM = Black Lives Matter. "Anti-Conservative" represents a misaligned bias. "Anti-BLM" and "Anti-#MeToo" represent conflated biases.

## Studies E4a-b

In Studies E4a-b, we examine whether the mere existence of political difference (without an explicit statement) would reduce attributions of racism (E4a) and sexism (E4b).

### Manipulation and Measures

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Studies 2a/b in the main text). Participants were randomly assigned to one of three conditions: a "racial difference" control condition (E4a) or a "gender difference" control condition (E4b) (where only a racial difference or a gender difference was described), a "political difference" condition (where only a political difference was described), or a "both" condition (where both demographic and political differences were described). Those in the control condition read the following, with differences between Studies E4a/b in bold:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a **[White/male]** manager, and **[Alex/Sarah]**, a **[Black/female]** sales associate. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer, **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

Participants in the "political difference" treatment condition read:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a manager who everyone knows is a Republican, and **[Alex/Sarah]**, a sales associate who everyone knows is a Democrat. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer, **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

Finally, participants in the "both" treatment condition read:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a **[White/male]** manager who everyone knows is a Republican, and **[Alex/Sarah]**, a **[Black/female]** sales associate who everyone knows is a Democrat. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer, **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

***Racism/Sexism Perceptions (Studies E4a/4b)***. We assessed perceptions of the actor's racism/sexism using the same items as Studies 2a-b in the main text (α = .98/.96).

**Results**

As preregistered, using ANOVA with Bonferroni multiple comparison tests, we found differences across conditions for both racism (E4a: $F(2, 529) = 62.48$, $p < .001$) and sexism (E4b: $F(2, 525) = 3.65$, $p = .027$). We found that participants in Study E4a perceived less racism in the "political difference" treatment condition ($M = 2.8$, $SD = 1.0$) relative to the "racial difference" control ($M = 3.8$, $SD = 1.1$; $t(529) = 9.31$, $p < .001$, $d = 0.40$, 95% CI [0.32, 0.49]). This comparison is not surprising given that there was no mention of race in the political difference condition in this study; therefore, participants perceived more racism when the racial difference was mentioned.

More importantly, we observed no differences in perceptions of racism between the "racial difference" control and the treatment condition where "both" political and racial differences were described ($M = 3.9$, $SD = 1.0$; $t(529) = 0.70$, $p = 1.000$, $d = -0.03$, 95% CI [-0.12, 0.05]). That is, in the presence of a racial difference, mere knowledge of a political difference (without the actor making an explicit statement about politics) did not reduce perceptions of racism.

In Study E4b, we found no difference in perceived sexism between the "gender difference" control ($M = 4.4$, $SD = 0.8$) and the "political difference" treatment condition ($M = 4.3$, $SD = 0.9$; $t(525) = 1.39$, $p = .494$, $d = 0.06$, 95% CI [-0.02, 0.15]). Although these results differ from Study E4a, they are also not surprising because both conditions included names that identified a gender difference. Moreover, the manager in the scenario used the phrase "women like you" in both conditions, which may have prompted sexism perceptions with or without describing a gender difference (see Study E6, where we compare effects of gendered vs. gender-neutral language).

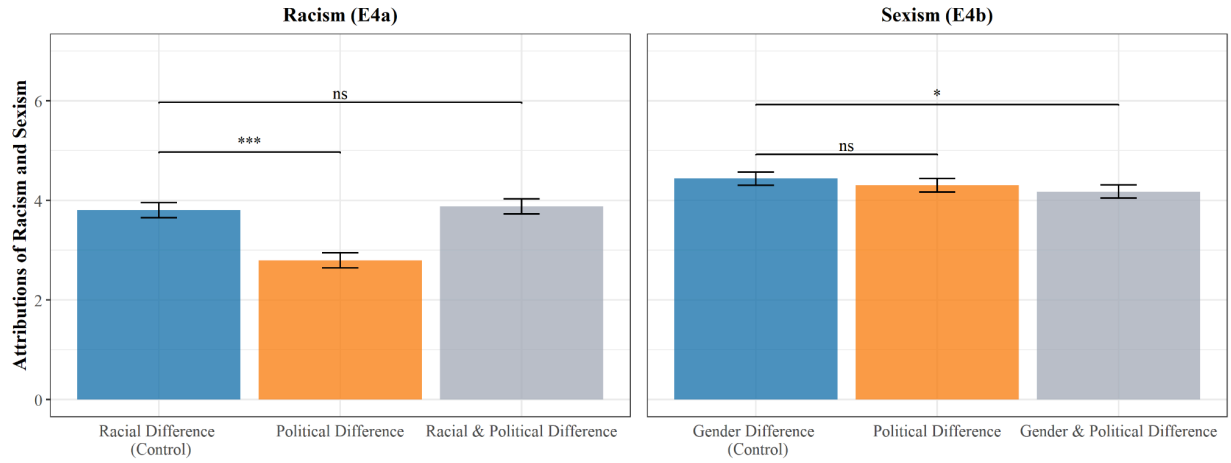
We did, however, find that, compared to the "gender difference" control, perceptions of sexism were lower in the treatment condition where "both" political and gender differences were described ($M = 4.2$, $SD = 1.0$; $t(525) = 2.70$, $p = .021$, $d = 0.12$, 95% CI [0.03, 0.20]). Thus, we find inconsistent results across these two studies. While the mere knowledge of political

difference (without an explicit political statement) was not sufficient to reduce perceptions of racism, it appeared sufficient to reduce perceptions of sexism. See Figure SM-K4.

Note that, while we considered many different ways to test this idea, each study design ultimately came with issues regarding the likelihood that participants would assume "default" characteristics, such as the target being White or a man. Thus, we preregistered the design that was most similar to our primary studies so that we could build on those studies in the clearest manner. Nonetheless, doing so meant that the political difference treatment condition in Study E4b inherently included references to gender (i.e., the target name being "Sarah" and the actor referring to "women like you").

**Figure SM-K4**

*Effects of the Mere Existence of Political Difference (Studies E4a-b)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.

## Studies E5a-b

In Studies E5a-b, we examine whether an actor expressing anti-liberal bias *without* the use of an anti-liberal pejorative would similarly divert third-party attributions of the actor's racism (E5a) and sexism (E5b).

### Manipulation and Measures

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Studies 2a/b in the main text). Participants were randomly assigned to one of three conditions. Those in the control condition read the following, with differences between Studies E5a/b in bold:

> Imagine that one day at work, you witnessed an intense conversation between **[Sam/John]**, a **[White/male]** manager, and **[Alex/Sarah]**, a **[Black/female]** sales associate. Angered by **[Alex/Sarah]**'s decision to refund money back to a customer,

> **[Sam/John]** told **[Alex/Sarah]** that he did not understand "why the firm keeps hiring **[people/women]** like you."

Participants in the treatment conditions read the same scenario with the exception of the quote in the last sentence, which randomized whether it included "liberal" or "woke liberal" before the words "[people/women] like you."

***Racism/Sexism Perceptions (Studies E5a/5b)***. We assessed perceptions of the actor's racism/sexism using the same items as Studies 2a-b in the main text (α = .97/.97).

**Results**

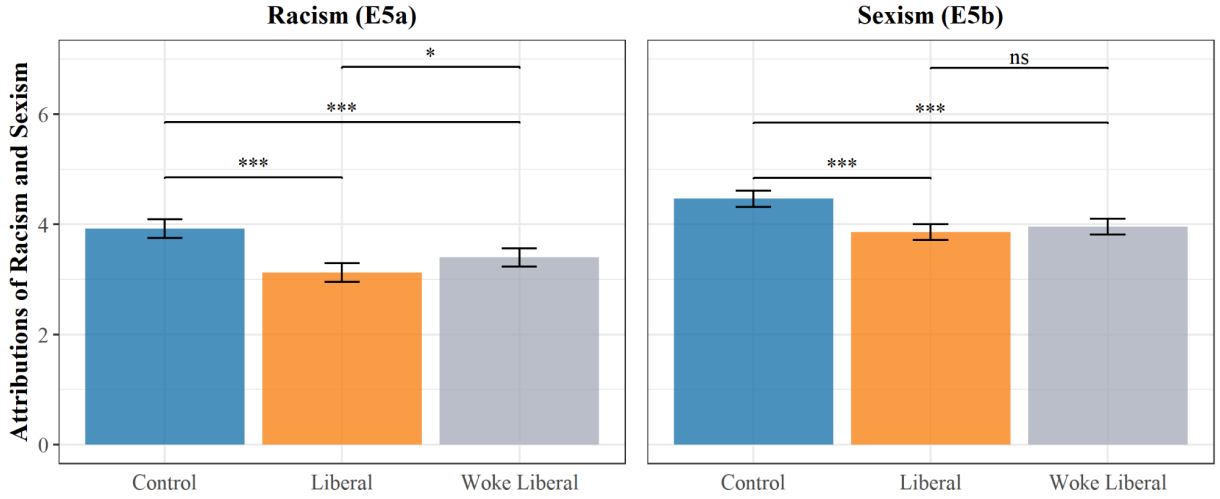
As preregistered, using ANOVA with Bonferroni multiple comparison tests, we again found differences across conditions for both racism (E5a: $F(2, 525) = 22.5$, $p < .001$) and sexism (E5b: $F(2, 527) = 19.45$, $p < .001$). Compared to the control ($M = 3.9$, $SD = 1.2$), participants in Study E5a perceived less racism regardless of whether an actor referred to a target as a "liberal" ($M = 3.1$, $SD = 1.1$; $t(525) = 6.60$, $p < .001$, $d = 0.29$, 95% CI [0.20, 0.38]) or as a "woke liberal" ($M = 3.4$, $SD = 1.1$; $t(525) = 4.33$, $p < .001$, $d = 0.19$, 95% CI [0.10, 0.28]). The effect of the "liberal" treatment was actually stronger than the effect of the "woke liberal" treatment ($F(1, 525) = 5.17$, $p = .023$).

The pattern of findings was similar in Study 5b: Relative to those in the control ($M = 4.5$, $SD = 0.8$), participants perceived less sexism regardless of whether an actor referred to a target as a "liberal" ($M = 3.9$, $SD = 1.1$; $t(527) = 5.82$, $p < .001$, $d = 0.25$, 95% CI [0.17, 0.34]) or as a "woke liberal" ($M = 4.0$, $SD = 1.1$; $t(527) = 4.86$, $p < .001$, $d = 0.21$, 95% CI [0.13, 0.30]). These two effects were statistically indistinguishable ($F(1, 525) = 0.90$, $p = .342$).

These results suggest that expressing anti-liberal bias—whether or not it includes a pejorative—reduces perceptions of racism and sexism. See Figure SM-K5.

**Figure SM-K5**

*Effects of Expressing Political Difference With and Without a Pejorative (Studies E5a-b)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.

## Study E6

In Study E6, we examine whether political plausible deniability—in the form of reduced third-party perceptions of sexism—depends on the actor using the gendered phrase "women like you" (vs. the gender-neutral phrase "people like you").

### Manipulation and Measures

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Study 2b in the main text). Participants were randomly assigned to one of four conditions, two of which served as control conditions. Those in the control conditions read the following (with differences between the two control conditions in bold):

> Imagine that one day at work, you witnessed an intense conversation between John, a male manager, and Sarah, a female sales associate. Angered by Sarah's decision to refund money back to a customer, John told Sarah that he did not understand "why the firm keeps hiring **[people/women]** like you."

Participants in the two treatment conditions read the same scenario with the exception of the quote in the last sentence, which included "liberal" before the words "[people/women] like you."

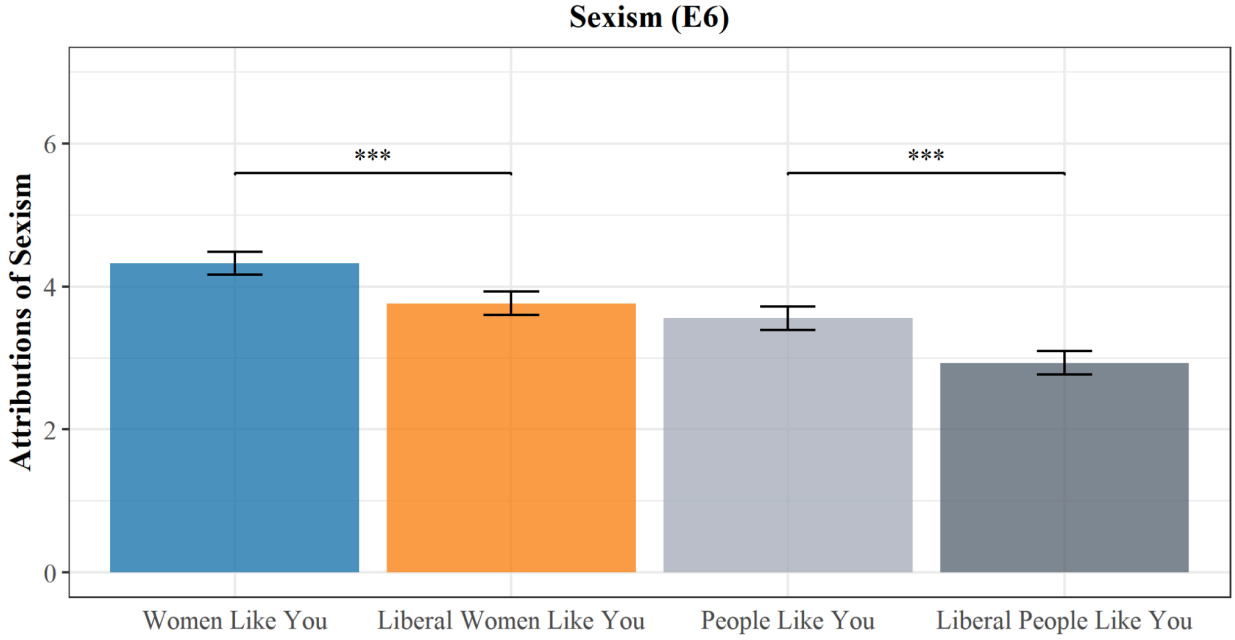
***Sexism Perceptions***. We assessed perceptions of the actor's sexism using the same items as Study 2b in the main text ($\alpha = .98$).

### Results

First, as preregistered, using ANOVA with Bonferroni multiple comparison tests, we found that perceptions of sexism differ across conditions ($F(3, 700) = 49.32$, $p < .001$). Relative to participants in the "people like you" (i.e., gender-neutral) control ($M = 3.6$, $SD = 1.1$), those in the "liberal people like you" (i.e., gender-neutral) treatment condition perceived less sexism ($M = 2.9$, $SD = 1.1$; $t(700) = 5.37$, $p < .001$, $d = 0.20$, 95% CI [0.13, 0.28]). Similarly, compared to participants in the "women like you" (i.e., gendered) control ($M = 4.3$, $SD = 0.9$), participants in the "liberal women like you" (i.e., gendered) treatment condition also perceived less sexism ($M = 3.8$, $SD = 1.2$; $t(700) = 4.82$, $p < .001$, $d = 0.18$, 95% CI [0.11, 0.26]). Moreover, these effects were statistically indistinguishable ($F(1, 700) = 0.16$, $p = .687$). These results suggest that political plausible deniability reduces perceptions of sexism regardless of whether the actor explicitly references the target's gender or not (at least for targets who are women). See Figure SM-K6.

**Figure SM-K6**

*Effects of Expressing Political Difference Using Gendered vs. Gender-Neutral Language (Study E6)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.

## Study E7

In Study E7, we examine how political plausible deniability—measured as reduced third-party perceptions of racism and sexism—operates at the intersection of the target's race and gender. In particular, we assessed whether third-party attributions of racism differ depending on whether the target is a Black man or a Black woman, and whether attributions of sexism differ depending on whether the target is a Black woman or a White woman.

### Manipulation and Measures

***Independent Variable***. All participants read a scenario in which the actor (a manager) expressed anger toward the target (a sales associate) for refunding money to a customer (i.e., the medium-stakes context from Study 2a in the main text, which did not rely on gendered names or use gendered language). Participants were randomly assigned to one of eight conditions, four of which served as control conditions. Those in the control conditions read the following (with differences between the four control conditions in bold):

> Imagine that one day at work, you witnessed an intense conversation between Sam, a White male manager, and Alex, a **[Black female/Black male/White female/White male]** sales associate. Angered by Alex's decision to refund money back to a customer, Sam told Alex that he did not understand "why the firm keeps hiring people like you."

Participants in the four treatment conditions read the same scenario with the exception of the quote in the last sentence, which included "liberal" before the words "people like you."

***Racism/Sexism Perceptions***. We assessed perceptions of the actor's racism/sexism using the same items as Studies 2a-b in the main text (α = .98/.98).

**Results**

First, as preregistered, using ANOVA with Bonferroni multiple comparison tests, we found that both perceptions of racism ($F(7, 1403) = 56.24$, $p < .001$) and perceptions of sexism ($F(7, 1403) = 35.61$, $p < .001$), differed across conditions.

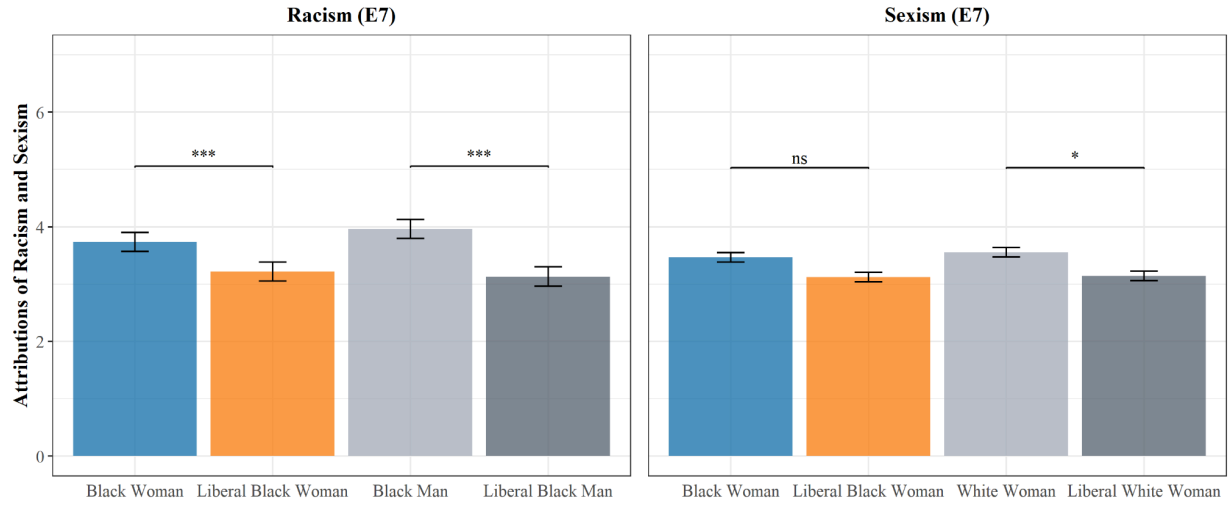
We found that political plausible deniability reduced perceived racism (against Black people) for both targets described as Black women and Black men. In particular, participants in the "anti-liberal" treatment conditions perceived less racism than those in the corresponding controls for scenarios with Black women (control: $M = 3.7$, $SD = 1.1$; treatment: $M = 3.2$, $SD = 1.3$; $t(1403) = 4.32$, $p < .001$, $d = 0.12$, 95% CI [0.06, 0.17]) and with Black men (control: $M = 4.0$, $SD = 1.0$; treatment: $M = 3.1$, $SD = 1.2$; $t(1403) = 6.91$, $p < .001$, $d = 0.18$, 95% CI [0.13, 0.24]).

We found that political plausibility reduced perceived sexism (against women); however, this effect holds only for targets described as White (but not Black) women. Specifically, participants in the "anti-liberal" treatment condition perceived less sexism than those in the corresponding control for scenarios with White women (control: $M = 3.6$, $SD = 1.1$; treatment: $M = 3.1$, $SD = 1.1$; $t(1403) = 3.58$, $p = .010$, $d = 0.10$, 95% CI [0.04, 0.15]). For scenarios with Black women, participants in the "anti-liberal" treatment condition perceived less sexism than participants in the control; however, this difference did not reach a conventional level of statistical significance (control: $M = 3.5$, $SD = 1.1$; treatment: $M = 3.1$, $SD = 1.1$; $t(1403) = 2.99$, $p = .081$, $d = 0.08$, 95% CI [0.03, 0.13]).

These results suggest that, while political plausible deniability effectively diverts attributions of racism across Black targets regardless of gender, its ability to reduce perceptions of sexism appears limited to cases where the target is a White (vs. Black) woman. The treatment did not reduce perceptions of bias against White people or bias against men for any race/gender combination. See Figure SM-K7.

**Figure SM-K7**

*Effects of Expressing Political Difference at the Intersection of Race and Gender (Study E7)*




*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$. ns = statistically nonsignificant. Capped bars represent 95% confidence intervals.